A STUDY OF SUBJECTIVE VIDEO QUALITY AT VARIOUS FRAME RATES

Alex Mackin, Fan Zhang and David R. Bull

Department of Electrical and Electronic Engineering, University of Bristol, BS8 1UB, UK
{A.Mackin, Fan.Zhang, Dave.Bull}@bristol.ac.uk

ABSTRACT

This paper presents a new video database (BVI-HFR), which contains content with a variety of frame rates from 15Hz to 120Hz, that can be used to demonstrate the benefits and limitations of higher frame rates, as well as investigating the role that frame rates play from capture to delivery. A characterization of the video database using low-level descriptors is also provided, which establishes that it successfully spans a variety of scene types and motions, and compares well to existing video databases. Subjective evaluations performed on the video database, have demonstrated a significant relationship between frame rates and perceived quality, up to 120Hz. They also confirm that the relationship between frame rate and perceived quality is content dependent.

***Index Terms*—** High frame rates, video database, subjective quality assessment.

1. INTRODUCTION

As the demand for higher quality and more immersive video content increases, the need to extend the current video parameter space of spatial resolutions and display sizes, to include, among other things, a wider color gamut, higher dynamic range and higher frame rates, becomes ever greater. The use of increased frame rate can provide a more realistic portrayal of a scene through a reduction in motion blur, while also minimizing temporal aliasing, and the associated visual artifacts. Higher frame rates also allow directors to have greater control and flexibility over the 'look' of the content, and offer easier conversion between formats [1].

The frame rate required to eliminate perceptible temporal aliasing depends on the retinal image velocity (deg/s) of a stimulus, as well as a number of others factors, including: eye movements [3], the luminance of the display [4, 5], color [6], viewing environment [7] and the distance from fixation [8]. However we can never sample at a high enough rate to capture optical reality [2], as there is no upper bound on the speed of a stimulus. We must also take into account the motion blur imposed by the camera shutter when considering frame rates, as temporal aliasing and motion blur are interdependent [9].

The authors acknowledge funding and support from BBC Research and Development, and EPSRC grant EP/J019291/1.

Notable research into the relationship between frame rate and perceived quality include the work by Ou *et al.* [10], Sugawara *et al.* [11] and Emoto *et al.* [12], but these have all been conducted based on different methodologies. Ou *et al.* consider frame rates only up to 30Hz, and while Emoto *et al.* consider frame rates up to 240Hz, they do not provide their source sequences. Sugawara *et al.* investigate the perception of motion blur, which is of key significance when analyzing frame rates. However there is little research that reports on how the video content affects the perception of quality at a given frame rate. Furthermore very few high frame rate (HFR) databases have been publicly released [13].

In this paper a new publicly available video database is presented, which spans a variety of scenes over a range of frame rates, from 15Hz to 120Hz. Alongside this we quantify the relationship between frame rate and perceptual quality, as well as assessing the content dependence of frame rates with respect to perceived quality. Our video database will enable others to investigate the role that frame rates play from capture to delivery, especially in the context of objective video quality metrics and video compression.

The rest of this paper is organized as follows: Section 2 describes the acquisition and characterization of the BVI-HFR video database, Section 3 outlines the experimental setup and methodology of our subjective experiments, the results of which are presented in Section 4. Conclusions are given in Section 5, along with suggestions for future work.

2. DATABASE DESCRIPTION

The latest video format standard, ITU-R Rec. 2020 [14], defines a wider color gamut, a higher dynamic range and an increased spatial resolution over previous video standards [15], it also supports frame rates up to 120Hz. Despite this, there is currently very little publicly available content shot at this upper frame rate limit. Providing video sequences at a variety of frame rates, up to 120Hz, will enable a thorough characterization of the benefits of higher frame rate material, as well as highlighting some of the challenges faced at higher frame rates, such as noticeable lighting variations between frames (caused by artificial lighting), higher compression rates, and more noise from the camera sensor (due to there being inherently less light within each exposure).

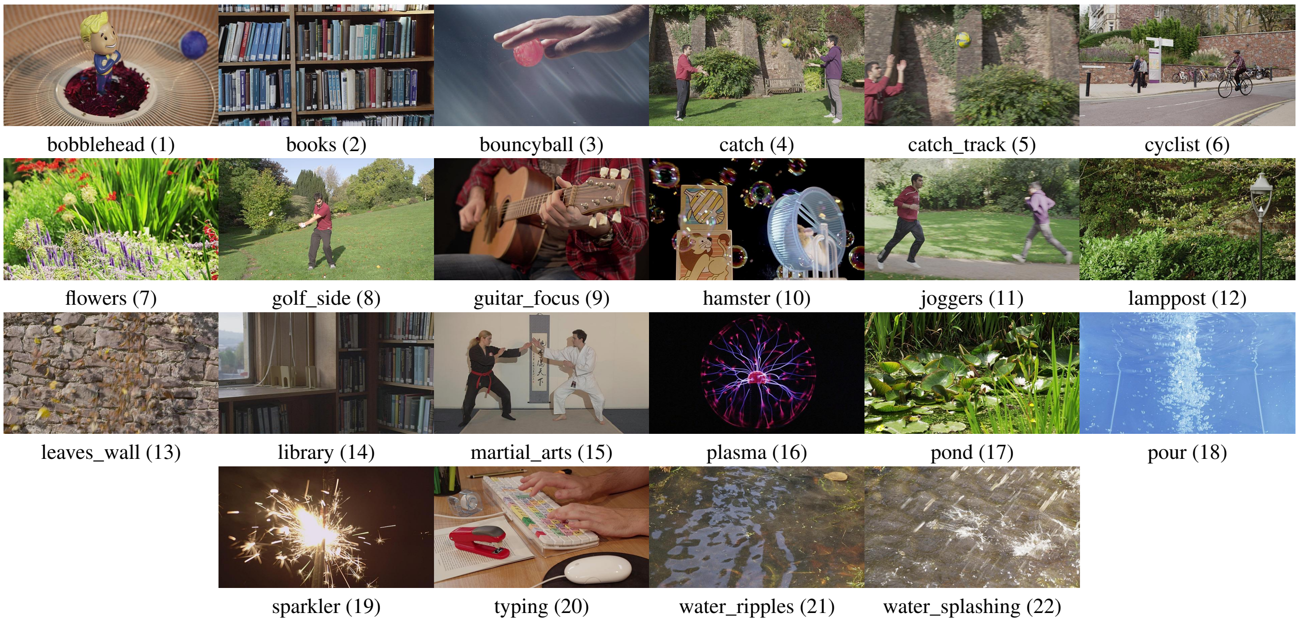

Fig. 1: Sample frames from the source sequences (120Hz) in the BVI-HFR video database along with their associated index.

2.1. Source sequences

The Bristol Vision Institute High Frame Rate (BVI-HFR) video database[1] contains 22 unique source sequences, which were all captured natively using a RED Epic-X video camera at a 3840×2160p (4K UHD) spatial resolution and a frame rate of 120Hz using a 360° shutter angle. These sequences were spatially down-sampled to 1920×1080p (HD) resolution using REDCINE-X software, into YUV 4:2:0 format. All sequences are 10 seconds in duration, and contain no shot transitions or audio components. The names, and associated index of each source sequence are shown in Fig. 1. There are four versions of each source sequence in the video database (88 in total), the original (120Hz) along with three temporally down-sampled versions (60, 30 and 15Hz) generated by averaging frames.

2.2. Content description and coverage

Three low-level descriptors were computed for each source sequence (120Hz) to characterize the content of the database: (i) spatial information (SI), which is an estimator of the amount of edge energy within each frame, (ii) temporal information (TI), which estimates the amount of motion energy between frames, and (iii) colorfulness (CF), which quantifies the variety, and intensity of colors within the scene. SI and CF are computed using the method described by Winkler [13], whereas TI is the mean difference in luma between frames:

$$\mathrm{TI} = \frac{1}{N-1} \sum_{t=1}^{N-1} \sum_{i,j}^{P} \frac{|I(i,j,t+1) - I(i,j,t)|}{P} \quad (1)$$

[1] http://data.bris.ac.uk/data/dataset/k8bfn0qsj9fs1rwnc2x75z6t7

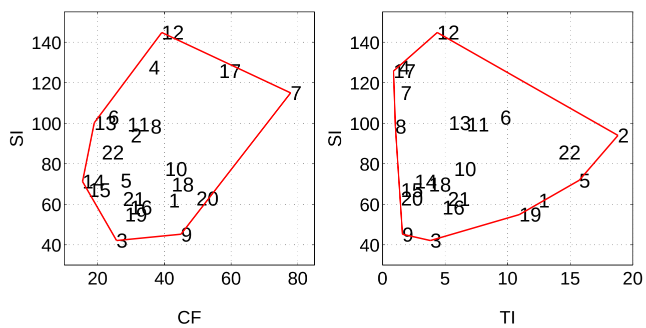

Fig. 2: Coverage of the source sequences: (left) SI vs. CF, (right) SI vs. TI.

where $I(i,j,t)$ is the luma value for pixel i,j in frame t, P is the total pixels per frame, and N is the number of frames.

Table 1 shows the range and uniformity characteristics of these three descriptors[2] alongside Winkler's MV metric [13] for the BVI-HFR source sequences (120Hz), where the coverage over the descriptors for the source sequences is shown in Fig. 2 (the points correspond to the indexes in Fig. 1). The BVI-HFR source sequences offer good coverage over the descriptors [13], with the relative range of SI being excellent. The relative total coverage of the source sequences is 0.41, which is slightly less than comparable databases with the same number of sequences [13]. This is because it is difficult to capture sequences that have both high MV and SI values, due to the increased blurring imposed by having a 360° shutter angle. Nonetheless these results indicate that the database successfully spans a variety of scenes and motions.

[2] As TI is calculated using a different method compared to Winkler we did not compute its relative range, as we have not considered any other databases.

Table 1: Characterization of the BVI-HFR source sequences (120Hz).

	SI	CF	MV	TI
Range	102.7	62.4	2.1	17.9
Relative range	0.68	0.62	0.7	-
Uniformity of coverage	0.85	0.84	0.86	0.84

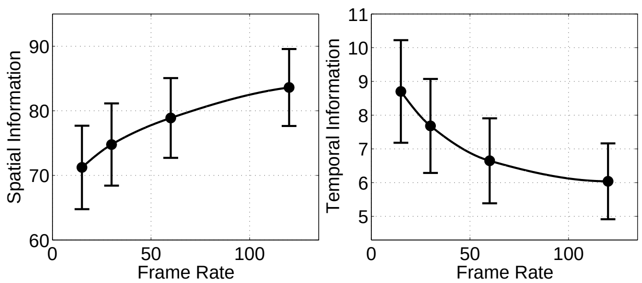

Fig. 3: The relationship between the descriptors (averaged over all relevant sequences) and frame rate in the BVI-HFR video database: (left) SI vs. Frame Rate, (right) TI vs. Frame Rate. Error bars represent standard error of the mean.

Fig. 3 shows how spatial and temporal information correlate with frame rate across the video database. At higher frame rates there is increased edge energy due to the reduction in motion blur, shown by the increase in spatial information. Also, as expected, the amount of motion energy between frames decreases with increased frame rates. These changes in the characteristic of the video signal at higher frame rates will have consequences with respect to compression, as at higher frame rates, higher frequency edges and reduced motion blur will influence both the transform and quantization stages. This along with a reduction in motion energy between frames should lead to smaller and more accurate motion vectors. This information could be used to improve compressive performance at higher frame rates, for example by exploiting a longer Group of Picture (GOP) structure.

3. SUBJECTIVE EVALUATIONS

3.1. Experimental setup

All sequences were displayed on a calibrated BenQ XL2720Z LCD monitor, with a peak luminance of 200 cd/m^2. A spatial resolution of 1920×1080 was used, measuring 598×336 mm, along with a static contrast ratio of 1000:1, and a refresh rate of 120Hz. The display was connected to a Windows PC running Matlab R2013a and Psychtoolbox 3.0. The viewing distance was set to be 1.68m (5H), which is calculated to be within the optimal range based on the spatial acuity of the retina [16, 17]. The environmental conditions conformed to the laboratory conditions outlined in ITU-R BT. 500-13 [18].

Prior to testing, each participant was given instructions related to the testing process. A complete session for each experiment lasted no longer than 30 minutes.

Each trial consisted of the participant viewing a 3s mid-level gray screen, before viewing a randomly chosen sequence, followed by a 3s mid-level gray screen in which they record their subjective quality score using the single stimulus continuous quality evaluation (SSCQE) method [18].

Mean opinion scores (MOS) were calculated for each test sequence by taking the average opinion score for all the participants, along with the standard error. The same observer screening procedure as outlined in [18] is used, resulting in none of the participants being removed in Experiment 1, and one participant being removed in Experiment 2.

3.2. Experiment 1: Temporal down-sampling

This experiment was designed to validate the temporal down-sampling methodology used in creating the BVI-HFR video database. Three versions of each source sequences outlined in Fig. 1 were generated, at frame rates of 60, 30 and 15Hz by frame averaging[3] (as used previously in [1]). To establish the degree of similarity between temporally down-sampled content, and content shot natively at that frame rate, we performed a subjective experiment comparing the two methods.

Twenty-one undergraduate and postgraduate students (15 male and 6 female) at the University of Bristol were paid to participate in the experiment. The average age of participants was 24.2 years, and all had normal or corrected-to-normal color vision, which was verified by the use of a Snellen chart. The experiment comprised three unique 3-second test sequences, *truck, exectoy and fan*, shot natively at both 30Hz and 60Hz. The same sequence was then shot at 120Hz, and temporally down-sampled to both 30Hz and 60Hz by averaging frames, giving a total of twelve sequences. All three unique sequences were highly repeatable, meaning that an identical scene was effectively captured at all frame rates.

3.3. Experiment 2: Quality assessment

This experiment was designed to quantify the relationship between frame rates and perceived quality, as well as the content dependence of frame rates with respect to perceived quality.

Twenty-nine undergraduate and postgraduate students (23 male and 6 female) from the University of Bristol were paid to participate in the experiment. The average age of participants was 24.8 years, and all had normal or corrected-to-normal color vision, which was verified by the use of a Snellen chart.

Participants were asked to rate their perceived quality of all 88 sequences in the BVI-HFR video database (the source sequences plus the three temporally down-sampled versions). Before each test session the participants took part in a brief training session of 5 randomly selected sequences to acclimatise themselves to the testing process. The sequences used in the training session were independent from those in the BVI-HFR video database, and spanned a range of frame rates.

[3] A 360° shutter angle must be used as otherwise we get ghosting artifacts.

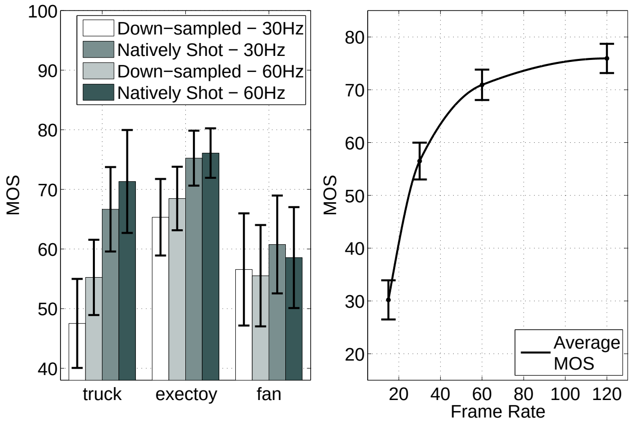

Fig. 4: (left) The results from Experiment 1. (right) The results from Experiment 2. Error bars represent standard error of the mean.

4. RESULTS AND DISCUSSION

An analysis of variance (ANOVA) [19] is used in this section, where all experimental data is verified for normality. F is the test statistic, with an assumed significance level of $p < 0.05$.

Fig. 4 (left) shows the difference in participant opinion scores from Experiment 1, between natively shot and temporally down-sampled sequences at 30Hz and 60Hz. A three-way repeated measures ANOVA (factor 1: frame rate, factor 2: sequence, factor 3: temporally down-sampled or natively shot) on the participant opinion scores shows no significant difference between temporally down-sampled and natively shot sequences, $F(2,40)$=0.164, p=0.849. This validates our approach of generating lower frame rate sequences by frame averaging.

The results from Experiment 2 are shown in Fig. 4 (right), which quantifies the relationship between frame rate and perceived quality. A one-way repeated measures ANOVA on the participant MOS scores for all sequences at various frame rates shows that the effect of frame rate is statistically significant with respect to perceived quality, $F(1.225, 34.295)$=198.85, p=0 (degrees of freedom adjusted using Greenhouse-Geisser correction due to violation of sphericity assumption). We also see a significant difference between 60Hz and 120Hz, $F(1,28)$=41.95, p=0. The increase in perceived quality at 120Hz is most likely due to a reduction in both motion blur and temporal aliasing artefacts. Improvement in quality beyond 120Hz is predicted for certain content, as the faster the speed of a stimulus relative to the retina, the higher the frame rate we will need [2].

From Fig. 4 (right) we can see an effect of diminishing returns with respect to perceived quality and increasing frame rates. Although these is a significant difference between 60Hz and 120Hz, it is smaller than that between 30Hz and 60Hz. These results may have implications when choosing the frame rate for content when bandwidth is limited.

Table 2: The results from the one-way repeated measures ANOVA, with degrees of freedom $F(1,28)$, on participant opinion scores for each sequence comparing 60Hz and 120Hz (sequences in **bold** are significant). The sequence # corresponds to the associated sequence index in Fig. 1.

#	1	**2**	3	**4**	**5**	**6**	7	**8**	9	**10**	**11**
F	.9	**33**	1.5	**15**	**15**	**5**	.6	**6**	.9	**10**	**6**
p	.3	**0**	.2	**0**	**0**	**.04**	.5	**.02**	.3	**0**	**.02**
#	12	13	**14**	15	**16**	17	**18**	19	**20**	21	**22**
F	.6	.3	**8**	3	**7**	.1	**16**	.04	**7**	.4	**4**
p	.5	.6	**.01**	.1	**.01**	.7	**0**	.9	**.01**	.6	**.04**

Previous research suggests that the relationship between perceived quality of a sequence and its frame rate is content dependent [12], and related to the speed of a tracked object. A two-way repeated measures ANOVA (factor 1: frame rate, factor 2: sequence) on the participant opinion scores affirms that this content dependence exists, $F(15.47, 433.26)$=6.3, p=0 (degrees of freedom adjusted using Greenhouse-Geisser correction due to violation of sphericity assumption).

Table 2 shows the results from a one-way repeated measures ANOVA on the participants opinion scores comparing 60Hz and 120Hz for every unique sequence in the BVI-HFR video database. These results show that some sequences (highlighted in **bold**) will have clear benefits at 120Hz compared to 60Hz, while other sequences have no significant difference in perceived quality. The reason for this content dependence can depend on many factors such as: speed and size of stimuli in the scene, viewing pattern, viewing conditions, and the amount of blurring in the scene. Camera motion also appears to contribute to the content dependence, as all the sequences with camera motion (*2,5,6,11,14*) show a significant difference between 60Hz and 120Hz.

5. CONCLUSION

We have introduced a new video database containing sequences at a range of frame rates, up to 120Hz. The database has been characterized using low-level descriptors, showing that it is sufficiently varied and complex, and that it compares well to existing video databases. We have demonstrated a significant relationship between frame rate and quality, where the effects of diminishing returns can be observed as frame rates increase. We have also shown that the perception of quality at a given frame rate is content dependent, demonstrating that some content will see clear benefits at 120Hz compared to 60Hz, while other content will see no significant difference. Future work will focus on using the database to understand the role that frame rates play from capture to delivery, specifically in the context of compression and objective video quality metrics. Furthermore, there is scope to use low-level descriptors such as spatial and temporal information to model the relationship between frame rate and perceived quality. This may be used in future perception-based rate-quality optimization processes

6. REFERENCES

[1] M. Armstrong, D. Flynn, M. Hammond, S. Jolly, and R. Salmon, "High frame-rate television," *BBC Research White Paper 169*, September 2008.

[2] A. Watson, "High frame rates and human vision: a view through the window of visibility," *SMPTE Motion Imaging Journal*, vol. 122, no. 2, pp. 18–32, 2013.

[3] S. Daly, "Engineering observations from spatiovelocity and spatiotemporal visual models," in *Photonics West '98 Electronic Imaging*. International Society for Optics and Photonics, 1998, pp. 180–191.

[4] J. Rabin, "Luminance effects on visual acuity and small letter contrast sensitivity," *Optometry & Vision Science*, vol. 71, no. 11, pp. 685–688, 1994.

[5] C. Tyler and R. Hamer, "Analysis of visual modulation sensitivity. iv. validity of the ferry-porter law," *JOSA A*, vol. 7, no. 4, pp. 743–758, 1990.

[6] J. Knöll, P. Binda, M. Morrone, and F. Bremmer, "Spatiotemporal profile of peri-saccadic contrast sensitivity," *Journal of vision*, vol. 11, no. 14, pp. 15, 2011.

[7] J. Van Hateren, "Spatiotemporal contrast sensitivity of early vision," *Vision research*, vol. 33, no. 2, pp. 257–267, 1993.

[8] J. Koenderink, M. Bouman, A. de Mesquita, and S. Slappendel, "Perimetry of contrast detection thresholds of moving spatial sine wave patterns. ii. the far peripheral visual field (eccentricity 0–50)," *J. Opt. Soc. Am*, vol. 68, no. 6, 1978.

[9] K. C. Noland, "The application of sampling theory to television frame rate requirements," *BBC Research & Development White Paper 282*, 2014.

[10] Y. Ou, T. Liu, Z. Zhao, Z. Ma, and Y. Wang, "Modeling the impact of frame rate on perceptual quality of video," *City*, vol. 70, no. 80, pp. 90, 2008.

[11] M. Sugawara, K. Omura, M. Emoto, and Y. Nojiri, "Temporal sampling parameters and motion portrayal of television," in *SID*, 2009, vol. 9, pp. 1200–1203.

[12] M. Emoto, Y. Kusakabe, and M. Sugawara, "High-frame-rate motion picture quality and its independence of viewing distance," *Journal of Display Technology*, vol. 10, no. 8, pp. 635–641, 2014.

[13] S. Winkler, "Analysis of public image and video databases for quality assessment," *Selected Topics in Signal Processing, IEEE Journal of*, vol. 6, no. 6, pp. 616–625, Oct 2012.

[14] International Telecom Union, "Recommendation ITU-R BT. 2020-1, *Parameter Values for Ultra-High Definition Television Systems for Production and International Programme Exchange*," 2014.

[15] International Telecom Union, "Recommendation ITU-R BT. 709, *Basic Parameter Values for the HDTV Standard for the Studio and for International Programme Exchange*," 1990.

[16] K. Wang, "Influence of viewing conditions on subjective and objective video quality assessment," M.S. thesis, University of Bristol, 2014.

[17] D. Bull, *Communicating Pictures: A Course in Image and Video Coding*, Elsevier, 2014.

[18] International Telecom Union, "Recommendation ITU-R BT. 500-13, *Methodology for the subjective assessment of the quality of television pictures*," 2012.

[19] D. Howell, *Statistical methods for psychology*, Cengage Learning, 2012.